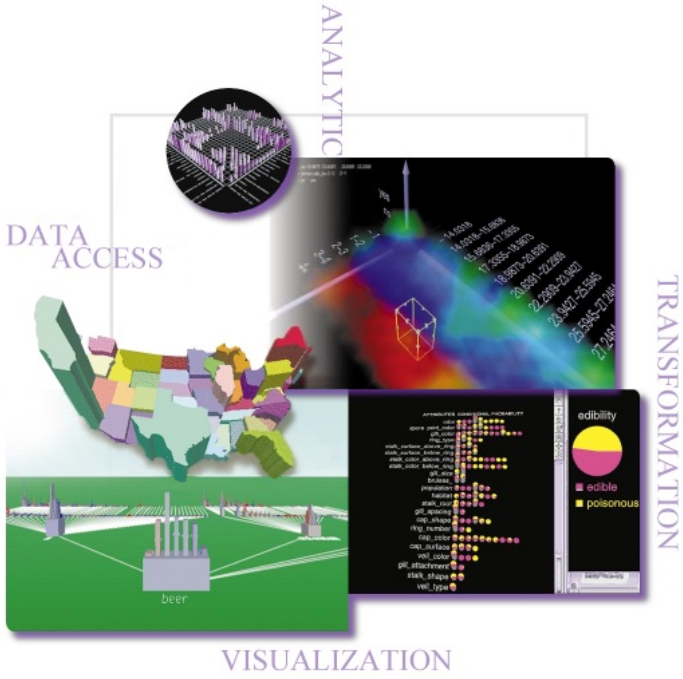
purple | insight

MineSet™ 3.2 Visual Data Mining Software

Features

- MineSet 100M, 500M on Microsoft Windows®
- MineSet Enterprise Server on Microsoft Windows®, and Silicon Graphics Inc (SGI) IRIX®
- MineSet Enterprise Client on Microsoft Windows®, and Silicon Graphics Inc (SGI) IRIX®
- Ability for Windows based clients to connect to both Windows and IRIX servers
- ODBC connectivity; provides access to SQL Server and other ODBC-compliant databases
- ActiveX visual components
- Software Development Kit providing an application programming interface to major MineSet components (available as an option)



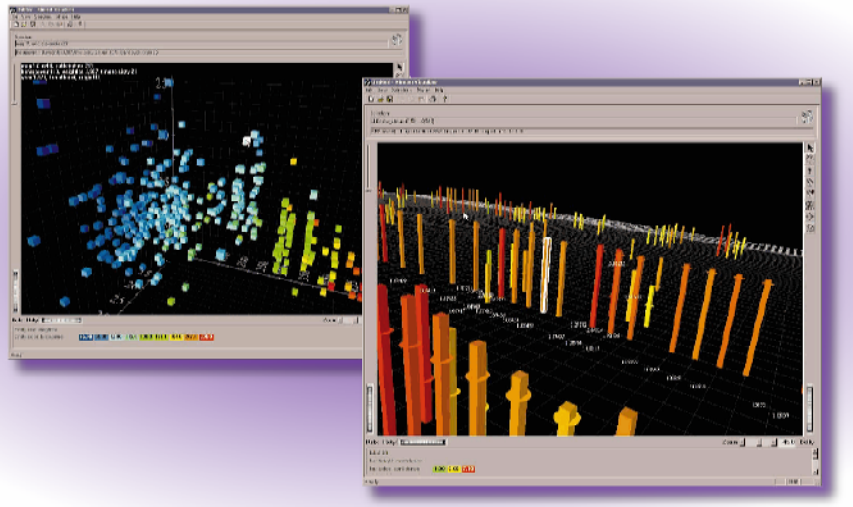
Buried within your vast store of transaction logs and legacy data, subtle correlations and elusive relationships hold the secrets to enhance your business. To reveal the hidden value in your data warehouse, Purple Insight MineSet is an integrated suite of software tools for data mining and data visualization. MineSet revolutionizes the way you use data mining for strategic business analysis. With the introduction of MineSet 3.2 , Purple Insight extends the power of visual data mining, previously available only to UNIX® users, to the Windows® desktop. By providing the same functionality on IRIX and Windows, Purple Insight allows business users to enjoy visual interpretation of complex data mining algorithms on PCs as well as take advantage of parallelization features on multiple CPU machines.

Intuitive Visual Data Mining Tools

MineSet enables interactive exploration of data through an advanced suite of visual tools for faster discovery of meaningful trends and relationships. The *Splat Visualizer* and the *Scatter Visualizer* represent complex data in up to eight dimensions. The *Map Visualizer* displays data with geographical relationships by using a map metaphor. Animation and view synchronization techniques are used to reveal patterns over critical dimensions such as time. The *Tree Visualizer* depicts data with hierarchical relationships utilizing a fly-through technique set in a 3D landscape. The *Statistics Visualizer* presents a visual summary of basic statistical information. Advanced drill-through techniques give you fast access to the original records that created entities within your visualization for additional analysis. The *Record Viewer* allows quick access to the original data with column sorting and HTML output capabilities.

Powerful Analytical Data Mining Algorithms with Matching Visualizers

Beneath the elegant interfaces of the visual tools, powerful analytic data mining algorithms build comprehensive models. MineSet includes multiple classifier inducers, including a *Decision Tree*, an *Option Tree*, *Evidence*, and *Decision Tables*. Boosting allows MineSet to further improve the accuracy of these classifiers. *Decision Tree* and *Option Tree* classification models are viewed through the *Tree Visualizer*. An



associated tool, the *Evidence Visualizer*, displays the structure and properties of the *Evidence Classifier* and supports instantaneous what-if analysis. The *Decision Table Visualizer* displays multidimensional data on important columns automatically chosen by the matching inducer. The visual-OLAP paradigm supports multilevel drill-through capability.

The *Regression Tree* inducer supports regression, and the resulting model is visualized using the *Tree Visualizer*. The *Association Rules Generator* analyzes data to discover product affinities and relationships between data entities. This feature can extract both one-to-one and multiway associations. The resulting one-to-one rules are depicted by the *Scatter Visualizer* in an easy-to-understand graphical format. Multiway rules are summarized in a text report. The *Column Importance* feature enables automatic or user-guided selection of important columns for use with the MineSet visualizers. The k-means-based Clustering algorithm is used for segmentation analysis and is visualized using the Cluster Visualizer.

Intuitive User Interface and Innovative Enterprise Solutions

MineSet is designed to fit the needs of users at several levels of technical ability. For technical users, MineSet offers unlimited potential with a complete suite of data mining tools, database integration, and scalable performance. To accommodate business users, data mining results and visualizations are easily deployed across corporate networks through point-and-click access. Analytical models developed in MineSet can be applied to any data set with the touch of a button. This easy scoring capability requires no programming or running of complicated batch programs. The MineSet Application Interface allows developers to embed MineSet tools within their customized solutions.

Integrating MineSet into Your Computing Environment

With seamless support for ODBC-compliant databases including SQL Server and DB2 and popular databases such as Oracle®, Informix®, and Sybase®, MineSet makes it easy to add new capabilities to your data warehouse. Queries to your relational database are supported through graphical or SQL commands. An open architecture allows MineSet to coexist with other data mining and visual tools, including SAS® software. Hot links help MineSet users share discoveries and results through Web pages, launching appropriate MineSet visualization tools for further analysis. The capability to connect to MineSet server on IRIX and Windows NT from a MineSet client on Windows provides unparalleled flexibility for business users who want to protect their investment in software as well as allow for explosive data growth.

Industry-Leading Scalability and Large Memory Support*

MineSet is optimized for the ultimate scalable performance on multiple CPU SGI machines. In addition to parallelization, MineSet offers very large scale memory support and data handling capability using the 64-bit implementation. MineSet allows business users unequalled flexibility and investment protection by providing the same back-end analytical processing capability on the popular Microsoft Windows Server® platform. Parallelization and 64-bit support are offered on the highly scalable SGI machines.

Using MineSet as a Platform for Developing Applications

MineSet 3.2 constitutes the ideal data mining application development platform through an advanced API and ActiveX visualization components. Business users and software vendors alike can reuse MineSet components or plug in custom algorithms.

Data Access and Transformation Features

- Support for ODBC-compliant databases, including SQL Server and DB2
- Support for access to Oracle, Informix, and Sybase running on all major UNIX platforms, including IBM, Sun, HP, and Digital
- Support for access to flat files (ASCII or binary)
- Data transformation history and graphical editing facility
- SAS file import/export utility
- Data transformation support for:
  - Automatic or user-specified binning of data
  - Data aggregations with indexed arrays using average, minimum, maximum, sum, and count
  - Data distribution (transpose)
  - New column creation using expressions
  - Data type conversion
  - Sampling
- Save and restore session management
- Client/server architecture
- Statistical tool for determining minimum, maximum, means, median, standard deviation, histograms, and quartiles

Analytical Data Mining Features

Classification
- Decision Tree
- Option Tree
- Evidence (simple Bayes)
- Decision Table

Associations
- One-to-one
- Multiway

Regression
- Regression trees

Clustering
- K-means
- Iterative k-means

Column Importance
- Automatic attribute selection

Advanced Analytical Options

User-specified parameters for control of classifier induction, including:
- Boosting
- Node-splitting criteria for decision tree induction
- Pruning factors for decision tree induction
- Laplace correction for evidence induction
- Automatic feature selection for evidence induction
- Data holdout percentage for automatic accuracy estimation (holdout and cross-validation)

Advanced features for classification, including:

- Loss matrices
- Lift curves
- Binning options
- Scoring
- Record weights
- Learning curves
- Backfitting

Visual Data Mining Features

- User-defined mapping of data to visualization components
- Visual and numeric normalization of data across visual components such as height, color, and size
- Discrete or continuous mapping of data to colors
- Web launching support
- Visual drill-up, drill-down, and drill-through to source data
- Animation of dependent data by up to two user-defined independent variables for trend analysis
- VCR-style animation playback
- Visual filtering and querying of data
- Synchronous animation of multiple visualizations
- 3D fly-through

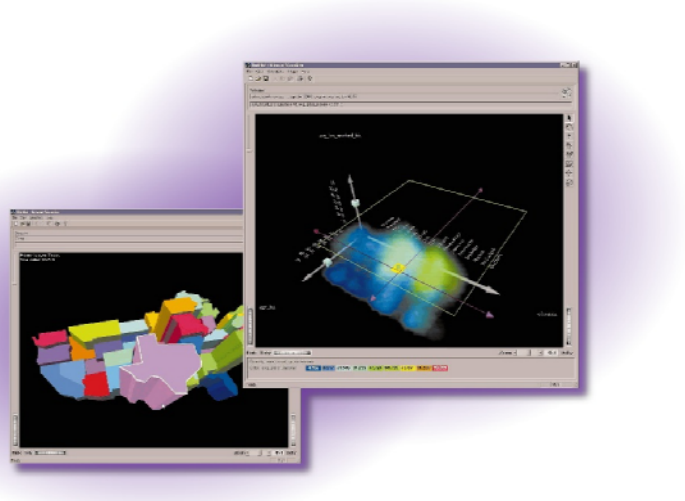
Integration with Windows Applications

Visualizers are provided as ActiveX components. This allows embedding of visualizations into Windows applications. Application vendors can embed industry-leading MineSet visualizations into business reports generated by their decision support tools.

Product Requirements

MineSet 3.2 100M, 500M, Enterprise Server and Enterprise Client are supported on all PC-compatible systems running Microsoft Windows® and Windows Server® platforms. Intel Pentium® processor or equivalent minimum. Memory requirements for the server vary with the size of the data. PC or server must have at least 64MB of memory; 96MB is recommended. Graphics accelerator cards enhance visualization performance; 1280x1024 resolution with 65K colors is the minimum recommended.

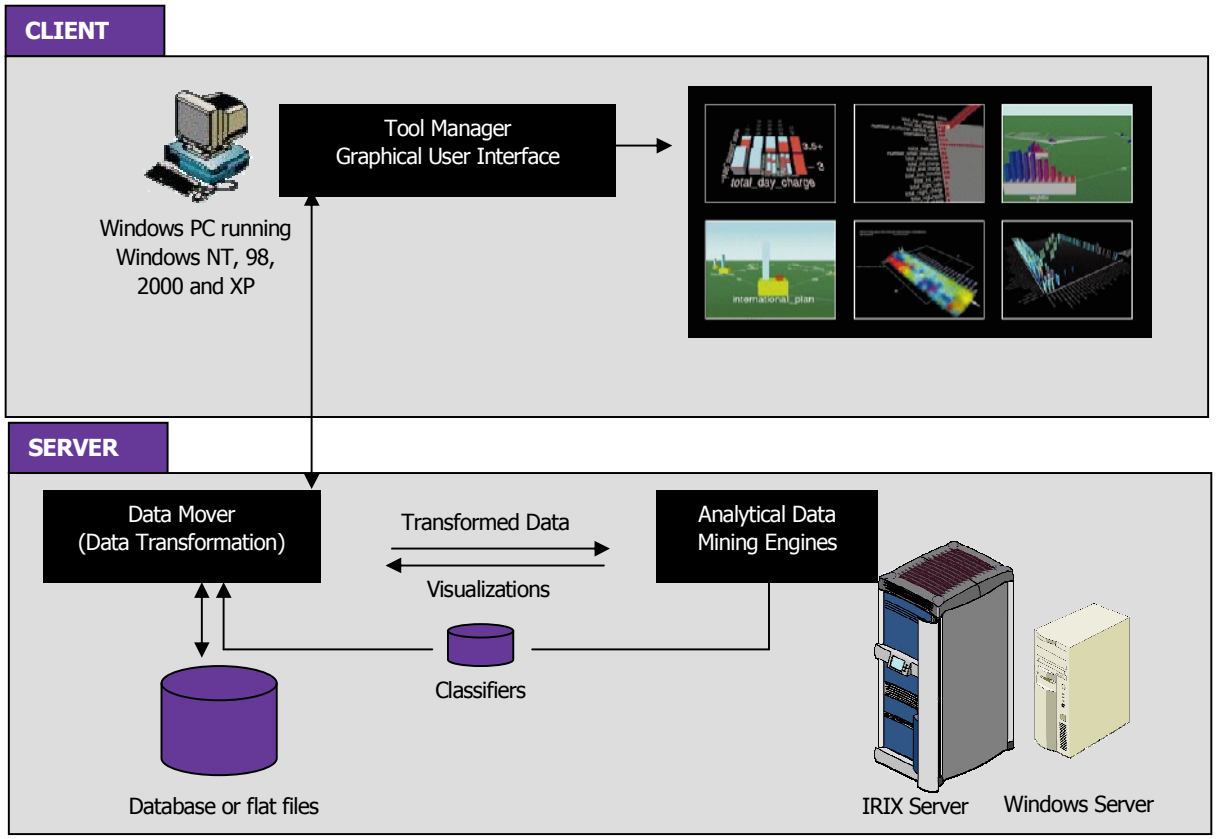
MineSet 3.2 Enterprise Server & Enterprise Client are available on Silicon Graphics Inc (SGI) platforms running IRIX® 6.4 and above.



MineSet System Architecture

MineSet employs a three-tiered architecture. The first tier is the client, which includes Tool Manager and the visualization tools. Tool Manager is the graphical interface through which the user interacts with MineSet. The visualization tools are used to display data and models of data generated by the mining algorithms. After invoking a visual tool with the Tool Manager, the user can interact directly with that tool and send information from it to other tools via Tool Manager. The second tier is the server, which includes Data Mover and the analytical mining engines. Data Mover is the database access and data transformation component of MineSet. It extracts data from the source, transforms it and orchestrates moving it from one MineSet component to another. The mining tools are used to generate models that can be applied to new data or visualized. The third tier is the data source, which includes the storage subsystem that maintains the users data. It can be either a file or a commercial database such as Oracle, Sybase or Informix. With MineSet Enterprise Software the tiers are not tied to specific machines: all three can reside on a single hardware platform or three separate machines.

Note: Each MineSet 100M and MineSet 500M software module may only be installed on one, single cpu machine and the data source may reside elsewhere.



This architecture provides the flexibility needed to scale with the size of the problem. Mining large databases requires powerful hardware: a substantial amount of memory, a number of fast CPUs and high I/O bandwidth. Desktop machines may not powerful enough to support data-mining data beyond several tens of megabytes. The MineSet architecture allows mining to be done on a powerful server machine, while the client can reside on a smaller PC or workstation. When working with small databases or during pilot projects, the architecture has the flexibility to allow the client, server and data source all to reside on a single machine.

Purple Insight Ltd
sales@purpleinsight.com
+44 (0)1452 813278
www.purpleinsight.com